

Dossier Next

Resumen. Dossier Next es una aplicación web que permite rastrear y clasificar de forma automática y periódica la información procedente de cualquier tipo de fuente de internet con contenido textual i.e. páginas web, periódicos digitales, boletines oficiales, etc., extrayendo únicamente la información más interesante para el usuario y descartando el resto. Para comenzar la recolección de información, el usuario debe indicar las fuentes a partir de las cuales desea obtener información, haciendo referencia a cada una las partes de interés para la recuperación de información, i.e. título, cuerpo, fecha, autor, etc. Para que el sistema pueda decidir qué contenido es relevante para el usuario, éste debe indicar también una serie de palabras claves, que deben aparecer en el documento en cuestión. Una vez establecida esta configuración del sistema se procede a descargar toda la información deseada por el usuario. Desde la página inicial se pueden visualizar los documentos por fecha y fuente.

Summary. Dossier Next is a web application that allows you to automatically and periodically track and classify information from any type of Internet source with textual content i.e. web pages, digital newspapers, official bulletins, etc., extracting only the most interesting information for the user and discarding the rest. To start collecting information, the user must define the sources from which to obtain information, indicating for each one the relevant document parts i.e. title, body, date, author, etc. How the user sets-up the searching criteria? The user must also indicate a series of keywords, which must appear in the document to be considered relevant. Once this system configuration is established, all the information desired by the user is downloaded. From the home page you can display the documents by date and source.

Especificaciones técnicas / Technical Specifications

Lenguaje de programación <i>Development Language</i>	Java, Javascript, HTML, CSS
Entorno Operativo <i>Operating Environment</i>	Linux, Windows, Mac
Versión <i>Version</i>	1.0

Requerimientos / Requirements

- **docker y docker-compose**
- Conexión a internet para la descarga de documentos
- 1GB de RAM o superior
- 300 MB de espacio en disco inicialmente, pero aumenta a medida que se descargan nuevos documentos

Instalación / Installation

- Descargar el fichero con la aplicación y descomprimirlo.
- Modificar el fichero **docker-compose.yml** con los datos necesarios.
- Ejecutar el siguiente comando:
./build.sh all

Ejecución / Run

- Dentro de la carpeta descargada, ejecutar el siguiente comando:
docker-compose up -d
- Para parar el servicio, ejecutar el siguiente comando:
docker-compose stop
- Para parar y destruir el servicio, ejecutar el siguiente comando:
docker-compose down

Dependencias / Dependencies

El listado completo de dependencias se encuentra en los archivos **core/pom.xml** , **api/pom.xml** y **web/package.json**. Aquí se incluye un listado de las dependencias que se descargarán de varios repositorios:

Dependencias Java / Java Dependencies · pom.xml

```
es.ua.sensei:sensei-core:jar:2.0.0
io.airlift:airline:jar:0.8
org.eclipse.persistence:eclipselink:jar:2.5.1
org.apache.lucene:lucene-snowball:jar:3.0.3
org.reflections:reflections:jar:0.9.11
com.fasterxml.jackson.core:jackson-databind:jar:2.5.3
org.mariadb.jdbc:mariadb-java-client:jar:1.2.0
org.jsoup:jsoup:jar:1.10.2
com.fasterxml.jackson.core:jackson-annotations:jar:2.5.0

com.fasterxml.jackson.jaxrs:jackson-jaxrs-json-provider:jar:2.5.3
commons-io:commons-io:jar:2.5
io.swagger:swagger-jersey2-jaxrs:jar:1.5.13
it.sauronsoftware.cron4j:cron4j:jar:2.2.5
org.glassfish.jersey.bundles:jaxrs-ri:jar:2.22
org.glassfish.jersey.media:jersey-media-multipart:jar:2.22
javax.ws.rs:javax.ws.rs-api:jar:2.0
```

Dependencias Node / Node Dependencies · package.json

```
angular:1.6.3
angular-cookies:1.6.3
angular-i18n:1.7.8
angular-loading-bar:0.9.0
angular-resource:1.6.3
angular-route:1.6.3
angular-sanitize:1.6.3
angular-ui-bootstrap:^2.5.6
angular-ui-codemirror:0.3.0
bootstrap:3.3.7
bootstrap-datepicker:1.8.0
bootstrap-daterangepicker:3.0.3
codemirror:5.47.0
moment:2.18.1
ng:0.0.0"
add-stream:^1.0.0

gulp:^3.9.1
gulp-angular-templatecache:^2.0.0
gulp-autoprefixer:^3.1.1
gulp-clean-css:^3.0.3
gulp-concat:^2.6.1
gulp-htmlmin:^3.0.0
gulp-if:^2.0.2
gulp-replace:^0.5.4
gulp-sass:^3.1.0
gulp-sftp:^0.1.5
gulp-sourcemaps:^2.4.1
gulp-sync:^0.1.4
gulp-uglify:^2.0.1
gulp-useref:^3.1.2
gulp-util:^3.0.8
gulp-watch:^4.3.11"
```

Estructura de ficheros / Files' structure

docker-compose.yml
README.md
build.sh
.gitignore
web/docker/nginx.conf
web/docker/entrypoint.sh
web/Dockerfile
web/README.md
web/.dockerignore
web/public/index.html
web/public/css/vendor.css
web/public/css/style.css
web/public/js/vendor.js
web/public/js/app.js
web/public/public.tgz
web/public/env.js
web/public/img/oportunity.png
web/public/img/oportunity2.png
web/public/img/dossierua.png
web/public/fonts/glyphicons-halflings-regular.woff
web/public/fonts/glyphicons-halflings-regular.eot
web/public/fonts/glyphicons-halflings-regular.woff2
web/public/fonts/glyphicons-halflings-regular.ttf
web/public/fonts/glyphicons-halflings-regular.svg
web/public/.htaccess
web/.gitignore
web/package-lock.json
web/package.json
web/gulpfile.js
web/src/index.html
web/src/app/home/home.html
web/src/app/home/home.js
web/src/app/home/home.module.js
web/src/app/_directives/directives.daterangepicker.js
web/src/app/_directives/directives.datepicker.js
web/src/app/_directives/directives.module.js
web/src/app/crawlers/crawlers.module.js
web/src/app/crawlers/delete.modal.html
web/src/app/crawlers/edit.modal.html
web/src/app/crawlers/list.html
web/src/app/crawlers/delete.modal.js
web/src/app/crawlers/list.js
web/src/app/crawlers/edit.modal.js
web/src/app/_common/topmenu.js
web/src/app/_common/topmenu.html
web/src/app/_common/common.module.js
web/src/app/_services/services.api.keyword.js
web/src/app/_services/services.api.source.js
web/src/app/_services/services.api.user.js
web/src/app/_services/services.api.crawler.js
web/src/app/_services/services.file.js
web/src/app/_services/services.utils.base64.js
web/src/app/_services/services.api.document.js
web/src/app/_services/services.array.js
web/src/app/_services/services.api.clipping.js
web/src/app/_services/services.api.category.js
web/src/app/_services/services.module.js
web/src/app/_services/services.api.authentication.js
web/src/app/users/role.modal.js
web/src/app/users/delete.modal.html
web/src/app/users/edit.modal.html
web/src/app/users/list.html
web/src/app/users/delete.modal.js
web/src/app/users/list.js
web/src/app/users/edit.modal.js
web/src/app/users/users.module.js
web/src/app/users/role.modal.html
web/src/app/sources/delete.modal.html
web/src/app/sources/edit.modal.html
web/src/app/sources/list.html
web/src/app/sources/sources.module.js
web/src/app/sources/delete.modal.js
web/src/app/sources/list.js
web/src/app/sources/edit.modal.js
web/src/app/_translate/messages.es.js
web/src/app/_translate/translate.filter.js
web/src/app/_translate/translate.module.js
web/src/app/documents/edit.modal.html
web/src/app/documents/documents.module.js
web/src/app/documents/list.html
web/src/app/documents/list.js
web/src/app/documents/edit.modal.js
web/src/app/documents/clipping.modal.js
web/src/app/documents/clipping.modal.html
web/src/app/app.module.js
web/src/app/keywords/delete.modal.html
web/src/app/keywords/edit.modal.html
web/src/app/keywords/list.html
web/src/app/keywords/delete.modal.js
web/src/app/keywords/list.js
web/src/app/keywords/edit.modal.js
web/src/app/keywords/keywords.module.js
web/src/app/app.routes.js
web/src/app/login/login.html
web/src/app/login/login.js
web/src/app/login/login.module.js
web/src/env.local.js
web/src/scss/style.scss
web/src/img/oportunity.png
web/src/img/oportunity2.png
web/src/img/dossierua.png
web/src/env.docker.js
web/src/.htaccess

core/.gitignore
core/pom.xml
core/README.md
core/log4j.properties
core/src/test/java/es/ua/oportunity/classifier/TestKeywordsSimpleClassifier.java
core/src/main/resources/META-INF/persistence.xml
core/src/main/resources/META-INF/persistence_1.xml
core/src/main/resources/oportunity.sql
core/src/main/resources/StopWords.txt
core/src/main/java/es/ua/oportunity/CommandExportComplex.java
core/src/main/java/es/ua/oportunity/entity/CrawlerParam.java
core/src/main/java/es/ua/oportunity/entity/Category.java
core/src/main/java/es/ua/oportunity/entity/User.java
core/src/main/java/es/ua/oportunity/entity/Document.java
core/src/main/java/es/ua/oportunity/entity/CrawlerParamId.java
core/src/main/java/es/ua/oportunity/entity/SourceParamId.java
core/src/main/java/es/ua/oportunity/entity/Taxonomy.java
core/src/main/java/es/ua/oportunity/entity/Clipping.java
core/src/main/java/es/ua/oportunity/entity/Crawler.java
core/src/main/java/es/ua/oportunity/entity/Source.java
core/src/main/java/es/ua/oportunity/entity/SourceParam.java
core/src/main/java/es/ua/oportunity/entity/Keyword.java
core/src/main/java/es/ua/oportunity/crawler/SenseiController.java
core/src/main/java/es/ua/oportunity/crawler/PyramidCrawler.java
core/src/main/java/es/ua/oportunity/crawler/CrawlerFuture.java
core/src/main/java/es/ua/oportunity/crawler/CrawlerException.java
core/src/main/java/es/ua/oportunity/crawler/ParamSpec.java
core/src/main/java/es/ua/oportunity/crawler/sensei/SenseiListenerNews.java
core/src/main/java/es/ua/oportunity/crawler/sensei/SenseiCrawler.java
core/src/main/java/es/ua/oportunity/crawler/sensei/SenseiListener.java
core/src/main/java/es/ua/oportunity/crawler/CrawlerPool.java
core/src/main/java/es/ua/oportunity/crawler/CrawlerType.java
core/src/main/java/es/ua/oportunity/controller/ControllerException.java
core/src/main/java/es/ua/oportunity/controller/Controller.java
core/src/main/java/es/ua/oportunity/controller/SenseiListener.java
core/src/main/java/es/ua/oportunity/classifier/KeywordsSimpleClassifier.java
core/src/main/java/es/ua/oportunity/classifier/SnowballUtil.java
core/src/main/java/es/ua/oportunity/classifier/Classifier.java
core/src/main/java/es/ua/oportunity/CommandImport.java
core/src/main/java/es/ua/oportunity/Main.java
core/src/main/java/es/ua/oportunity/service/SourceServiceJPA.java
core/src/main/java/es/ua/oportunity/service/FilteredList.java
core/src/main/java/es/ua/oportunity/service/ServiceException.java
core/src/main/java/es/ua/oportunity/service/UserServiceJPA.java
core/src/main/java/es/ua/oportunity/service/UserService.java
core/src/main/java/es/ua/oportunity/service/SourceService.java
core/src/main/java/es/ua/oportunity/service/CrawlerServiceJPA.java
core/src/main/java/es/ua/oportunity/service/CategoryServiceJPA.java
core/src/main/java/es/ua/oportunity/service/ServiceJPA.java
core/src/main/java/es/ua/oportunity/service/DocumentService.java
core/src/main/java/es/ua/oportunity/service/ClippingService.java
core/src/main/java/es/ua/oportunity/service/KeywordServiceJPA.java
core/src/main/java/es/ua/oportunity/service/DocumentServiceJPA.java
core/src/main/java/es/ua/oportunity/service/KeywordService.java
core/src/main/java/es/ua/oportunity/service/Service.java
core/src/main/java/es/ua/oportunity/service/InstallService.java
core/src/main/java/es/ua/oportunity/service/AuthenticationService.java
core/src/main/java/es/ua/oportunity/service/CategoryService.java
core/src/main/java/es/ua/oportunity/service/CrawlerService.java
core/src/main/java/es/ua/oportunity/service/TaxonomyService.java
core/src/main/java/es/ua/oportunity/service/TaxonomyServiceJPA.java
core/src/main/java/es/ua/oportunity/service/AuthenticationServiceJPA.java

core/src/main/java/es/ua/oportunity/service/ClippingServiceJPA.java
core/src/main/java/es/ua/oportunity/CommandExport.java
core/src/main/java/es/ua/oportunity/CommandStart.javaapi/Dockerfile
api/pom.xml
api/README.md
api/.gitignore
api/src/main/webapp/favicon-16x16.png
api/src/main/webapp/swagger-ui.js
api/src/main/webapp/swagger-ui-standalone-preset.js.map
api/src/main/webapp/swagger-ui.css
api/src/main/webapp/swagger-ui.js.map
api/src/main/webapp/swagger-ui-bundle.js
api/src/main/webapp/META-INF/context.xml
api/src/main/webapp/swagger-ui.css.map
api/src/main/webapp/swagger-ui-standalone-preset.js
api/src/main/webapp/WEB-INF/classes/META-INF/persistence.xml
api/src/main/webapp/WEB-INF/local.properties
api/src/main/webapp/WEB-INF/base.properties
api/src/main/webapp/WEB-INF/i18n/es.properties
api/src/main/webapp/WEB-INF/jsp/index.jsp
api/src/main/webapp/swagger-ui-bundle.js.map
api/src/main/webapp/oauth2-redirect.html
api/src/main/webapp/favicon-32x32.png
api/src/main/java/es/ua/oportunity/api/servlet/DocsServlet.java
api/src/main/java/es/ua/oportunity/api/OportunityAPIApplication.java
api/src/main/java/es/ua/oportunity/api/provider/KeywordServiceProvider.java
api/src/main/java/es/ua/oportunity/api/provider/UserServiceProvider.java
api/src/main/java/es/ua/oportunity/api/provider/CrawlerServiceProvider.java
api/src/main/java/es/ua/oportunity/api/provider/AuthenticationServiceProvider.java
api/src/main/java/es/ua/oportunity/api/provider/CategoryServiceProvider.java
api/src/main/java/es/ua/oportunity/api/provider/ClippingServiceProvider.java
api/src/main/java/es/ua/oportunity/api/provider/DocumentServiceProvider.java
api/src/main/java/es/ua/oportunity/api/provider/TaxonomyServiceProvider.java
api/src/main/java/es/ua/oportunity/api/provider/PropertiesProvider.java
api/src/main/java/es/ua/oportunity/api/provider/SourceServiceProvider.java
api/src/main/java/es/ua/oportunity/api/OportunityAPIListener.java
api/src/main/java/es/ua/oportunity/api/filter/CORSFilter.java
api/src/main/java/es/ua/oportunity/api/filter/AuthenticationFilter.java
api/src/main/java/es/ua/oportunity/api/resource/AuthenticationResource.java
api/src/main/java/es/ua/oportunity/api/resource/CrawlerResource.java
api/src/main/java/es/ua/oportunity/api/resource/DocumentResource.java
api/src/main/java/es/ua/oportunity/api/resource/PublicSourceResource.java
api/src/main/java/es/ua/oportunity/api/resource/CategoryResource.java
api/src/main/java/es/ua/oportunity/api/resource/PublicDocumentResource.java
api/src/main/java/es/ua/oportunity/api/resource/UserResource.java
api/src/main/java/es/ua/oportunity/api/resource/KeywordResource.java
api/src/main/java/es/ua/oportunity/api/resource/CrawlerTypeResource.java
api/src/main/java/es/ua/oportunity/api/resource/PublicClippingResource.java
api/src/main/java/es/ua/oportunity/api/resource/ClippingResource.java
api/src/main/java/es/ua/oportunity/api/resource/SourceResource.java
api/src/main/java/es/ua/oportunity/api/exception/IllegalArgumentExceptionHandler.java
api/src/main/java/es/ua/oportunity/api/exception/ParamExceptionHandler.java
api/src/main/java/es/ua/oportunity/api/exception/AuthorizationExceptionHandler.java
api/src/main/java/es/ua/oportunity/api/exception/AuthorizationExceptionHandler.java
api/src/main/java/es/ua/oportunity/api/exception/GenericExceptionHandler.java